

**PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING
DETEKSI MALWARE DENGAN Z-SCORE NORMALIZATION
HASIL TERBAIK PADA RANDOM FOREST**

LAPORAN NON-REGULER SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh

ZAKA GILANG RAMADHAN

21.83.0661

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING
DETEKSI MALWARE DENGAN Z-SCORE NORMALIZATION
HASIL TERBAIK PADA RANDOM FOREST**

LAPORAN NON-REGULER SCIENTIST
untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh
ZAKA GILANG RAMADHAN
21.83.0661

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING
DETEKSI MALWARE DENGAN Z-SCORE NORMALIZATION
HASIL TERBAIK PADA RANDOM FOREST

yang diperstasikan dan disusun oleh

Zaka Gilang Rumadban

21.83.0661

Tanggal, 15 Desember 2025

Dosen Pembimbing

Wahid Miftahul Ashari, S.Kom., M.T.

NIK 190302432

HALAMAN PENGESAHAN

SKRIPSI NON REGULER - SCIENTIST
PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING
DETEKSI MALWARE DENGAN Z-SCORE NORMALIZATION
HASIL TERBAIK PADA RANDOM FOREST

yang disusun dan diajukan oleh

Zaka Gilang Ramadhan

21.83.0661

Telah dipertahankan di depan Dewan Penguji
pada tanggal 16 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Jeki Kuswanto, S.Kom., M.Kom.
NIK. 190302456

Muhammad Kopravi, S.Kom., M.Eng.
NIK. 190302454

Wahid Miftahul Ashari, S.Kom., M.T.
NIK. 190302452

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 16 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriadi, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Zaka Gilang Ramadhan
NIM : 21.83.0661

Menyatakan bahwa Skripsi dengan judul berikut:

**PERBANDINGAN KINERJA ALGORITMA MACHINE LEARNING DETEKSI
MALWARE DENGAN Z-SCORE NORMALIZATION HASIL TERBAIK PADA
RANDOM FOREST**

Dosen Pembimbing : Wahid Miftahul Ashari, S.Kom., M.T.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 16 Desember 2025

Yang Menyatakan,

Zaka Gilang Ramadhan

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur ke hadirat Allah Subhanahu wa Ta'ala atas limpahan rahmat, karunia, serta kekuatan yang telah diberikan selama proses penulisan skripsi ini, izinkan penulis mempersembahkan karya sederhana ini kepada:

- Kedua orang tua tercinta, yang senantiasa menjadi sumber semangat dalam setiap langkah perjalanan penulis. Terima kasih atas kasih sayang, kesabaran, doa yang tiada henti, serta dukungan moral maupun materiil yang telah mengiringi setiap proses yang penulis jalani. Tanpa kalian, pencapaian ini tidak akan mungkin terwujud.
- Keluarga besar, yang selalu memberikan semangat, kepercayaan, dan doa terbaiknya. Kehadiran kalian menjadi motivasi tersendiri bagi penulis untuk terus melangkah maju dan menyelesaikan tanggung jawab akademik ini hingga tuntas.
- Rekan-rekan seperjuangan dan Heyek Gank, yang telah memberikan warna dalam setiap perjalanan selama masa kuliah. Terima kasih atas kebersamaan, bantuan, dukungan, serta semangat yang kalian tularkan saat proses penyusunan skripsi ini berlangsung.
- Seluruh pihak yang secara langsung maupun tidak langsung telah membantu penulis dalam menyelesaikan studi dan penulisan skripsi ini, baik melalui bimbingan, motivasi, maupun dukungan lainnya.

Skripsi ini adalah wujud kecil dari perjuangan panjang yang penuh proses, doa, dan pengorbanan. Semoga karya ini dapat menjadi bentuk balas budi yang membawa manfaat dan menjadi pijakan awal untuk langkah selanjutnya.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas berkat rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi ini dengan judul “Perbandingan Kinerja Algoritma Machine Learning Deteksi Malware dengan Z-Score Normalization Hasil Terbaik pada Random Forest” sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknik Komputer di Universitas Amikom Yogyakarta.

Penulisan skripsi ini tidak lepas dari bimbingan, dukungan, dan bantuan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

- Orang tua tercinta yang selalu memberikan doa, dukungan, dan kasih sayang yang tiada henti.
- Dosen pembimbing yang telah memberikan arahan, bimbingan, serta motivasi yang luar biasa selama proses penyusunan skripsi ini.
- Seluruh dosen dan staf akademik di lingkungan Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.
- Teman-teman Heyek Gank, yang telah memberikan dukungan moral, motivasi, dan kerja sama selama ini.
- Playlist musik JJ yang selalu menemani dalam setiap prosesnya.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, baik dari segi isi maupun penyusunannya. Oleh karena itu, penulis mengharapkan kritik dan saran yang konstruktif demi perbaikan dan pengembangan pengetahuan di masa depan. Semoga skripsi ini dapat memberikan kontribusi yang bermanfaat bagi perkembangan ilmu pengetahuan, khususnya di bidang cyber security.

Akhir kata, penulis berharap semoga skripsi ini dapat bermanfaat bagi para pembaca, terutama dalam praktis di bidang cyber security.

Yogyakarta, 15 Desember 2025

Penulis,

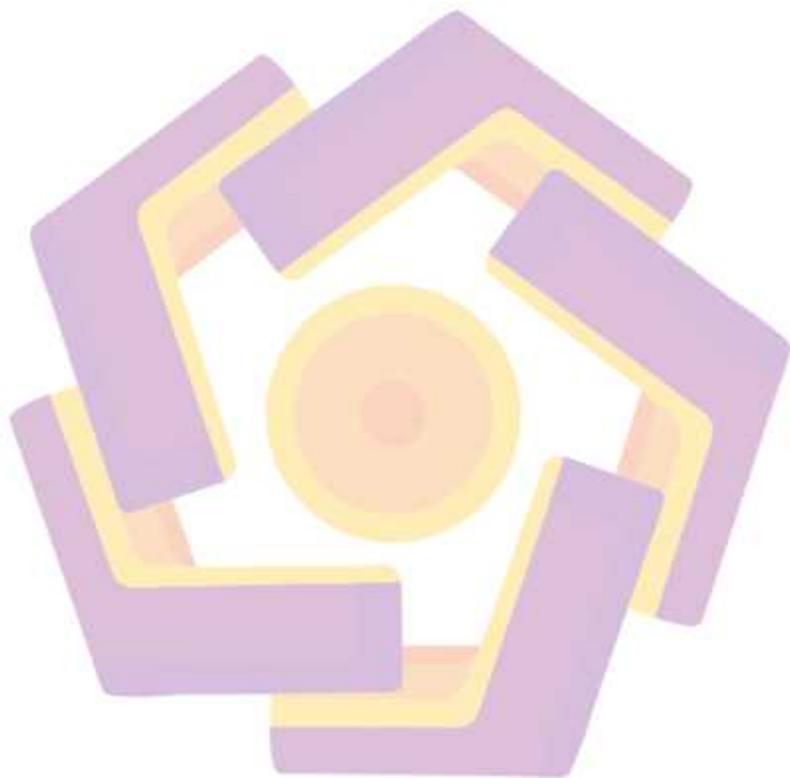
DAFTAR ISI

HALAMAN JUDUL	I
HALAMAN PERSETUJUAN	II
HALAMAN PENGESAHAN	III
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	IV
HALAMAN PERSEMBAHAN	V
KATA PENGANTAR.....	VI
DAFTAR ISI.....	VII
DAFTAR TABEL	IX
DAFTAR GAMBAR.....	X
DAFTAR LAMPIRAN	XI
INTISARI	XII
<i>ABSTRACT</i>	XIII
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	13
BAB III METODE PENELITIAN	24

3.1	Objek Penelitian.....	24
3.2	Alur Penelitian	24
3.2.1	Dataset.....	25
3.2.2	Data Pre-Processing	25
3.2.3	Application Of Algorithm.....	28
A.	Support Vector Machine	28
B.	XGBoost	28
C.	Random Forest.....	29
D.	Evaluation Matrix	29
3.3	Alat dan Bahan.....	30
BAB IV	HASIL DAN PEMBAHASAN	31
4.1	Distribusi Dataset.....	31
4.2	Confusion Matrix.....	33
4.3	Hasil Evaluasi Algoritma Machine Learning	34
BAB V	PENUTUP.....	38
5.1	Kesimpulan.....	38
5.2	Saran	38
REFERENSI	39	
LAMPIRAN	42	

DAFTAR TABEL

Table 2.1. Keaslian Penelitian	10
--------------------------------------	----

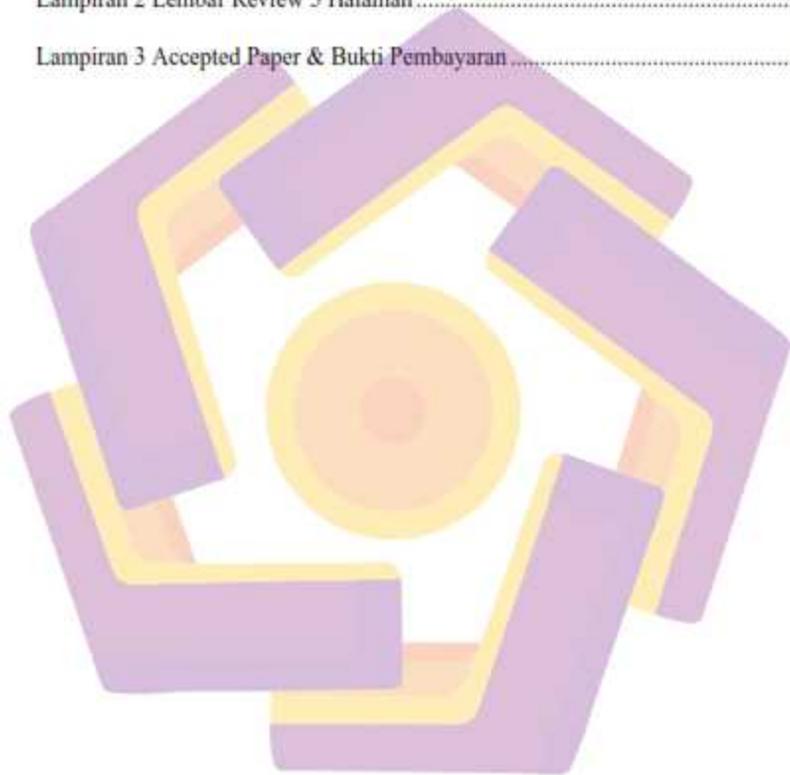


DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	24
Gambar 4.1 Perbandingan Distribusi Kelas.....	32
Gambar 4.2 Perbandingan Confusion Matrix pada Model Algoritma.....	33
Gambar 4.3 Perfoma Presisi, Recal, dan F1-Score pada Setiap Model Algoritma	34
Gambar 4.4 Perbandingan Akurasi Pada Model Algoritma	35
Gambar 4.5 Perbandingan ROC pada Model Algoritma.....	36
Gambar 4.6 Perbandingan K-Fold pada Model Algoritma.....	37

DAFTAR LAMPIRAN

Lampiran 1 Letter Of Acceptance (LOA).....	42
Lampiran 2 Lembar Review 5 Halaman.....	43
Lampiran 3 Accepted Paper & Bukti Pembayaran.....	48



INTISARI

Deteksi malware merupakan tantangan utama dalam dunia keamanan siber, terutama dengan meningkatnya kompleksitas dan variasi serangan. Pendekatan tradisional seringkali tidak mampu mengidentifikasi ancaman baru ini, sehingga machine learning (ML) menjadi solusi yang efektif. Tujuan dari penelitian ini untuk membandingkan kinerja tiga algoritma machine learning, yaitu Random Forest, XGBoost, dan Support Vector Machine (SVM), dalam mendeteksi malware pada dataset SOMLAP yang terdiri dari file eksekusi Windows. Pemrosesan data, penggunaan teknik SMOTE untuk ketidakseimbangan kelas, dan penilaian menggunakan metrik seperti akurasi, precision, recall, dan skor F1 adalah semua bagian dari metodologi penelitian. Penelitian ini juga menerapkan Z-Score Normalization untuk mengurangi pengaruh nilai-nilai ekstrem dalam data, yang membantu model menangani data dengan skala yang berbeda. Hasil penelitian memperlihatkan dimana Random Forest memiliki performa paling unggul dengan akurasi 99.16%, dilanjutkan oleh XGBoost dan SVM. Random Forest unggul dalam keseimbangan antara precision, recall, dan akurasi, menjadikannya algoritma yang paling efektif dalam mendeteksi malware. Penelitian ini menyarankan untuk pengembangan algoritma lebih lanjut dengan menggunakan teknik ensemble dan optimasi lainnya untuk meningkatkan akurasi deteksi malware di masa depan.

Kata kunci: Dataset SOMLAP, Deteksi malware, Machine Learning, Random Forest, Z-Score Normalization.

ABSTRACT

Malware detection is a major challenge in the world of cybersecurity, especially with the increasing complexity and variety of attacks. Traditional approaches are often unable to identify these new threats, making machine learning (ML) an effective solution. The purpose of this study is to compare the performance of three machine learning algorithms, namely Random Forest, XGBoost, and Support Vector Machine (SVM), in detecting malware in the SOMLAP dataset consisting of Windows executable files. Data processing, the use of the SMOTE technique for class imbalance, and assessment using metrics such as accuracy, precision, recall, and F1 score are all part of the research methodology. This study also applies Z-Score Normalization to reduce the influence of extreme values in the data, which helps the model handle data with different scales. The results show that Random Forest has the best performance with an accuracy of 99.16%, followed by XGBoost and SVM. Random Forest excels in the balance between precision, recall, and accuracy, making it the most effective algorithm for detecting malware. This study suggests further algorithm development using ensemble techniques and other optimizations to improve malware detection accuracy in the future.

Keyword: *SOMLAP dataset, Malware detection, Machine learning, Random forest, Z-score normalization.*